

An update on Scikit-learn

0.21 and beyond

Andreas Müller

Associate Research Scientist
Columbia University
Scikit-learn Technical Committee



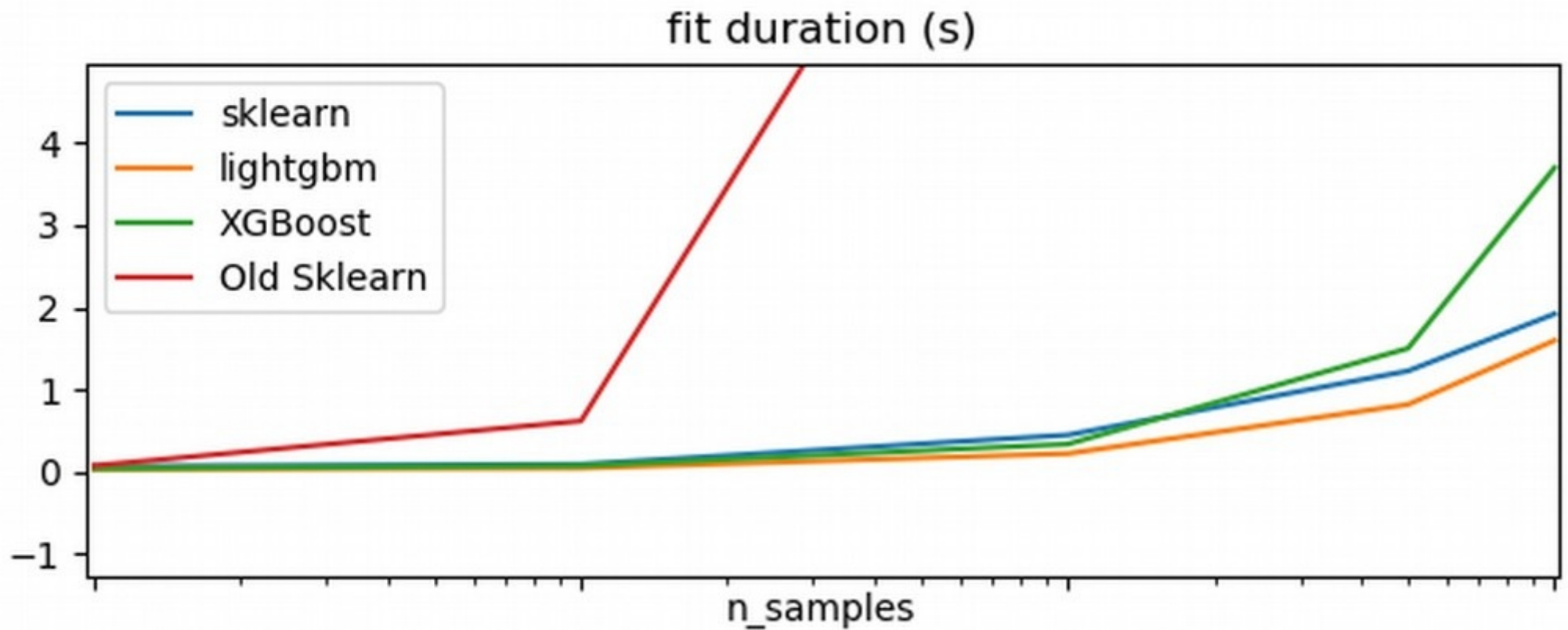
Alfred P. Sloan
FOUNDATION



COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

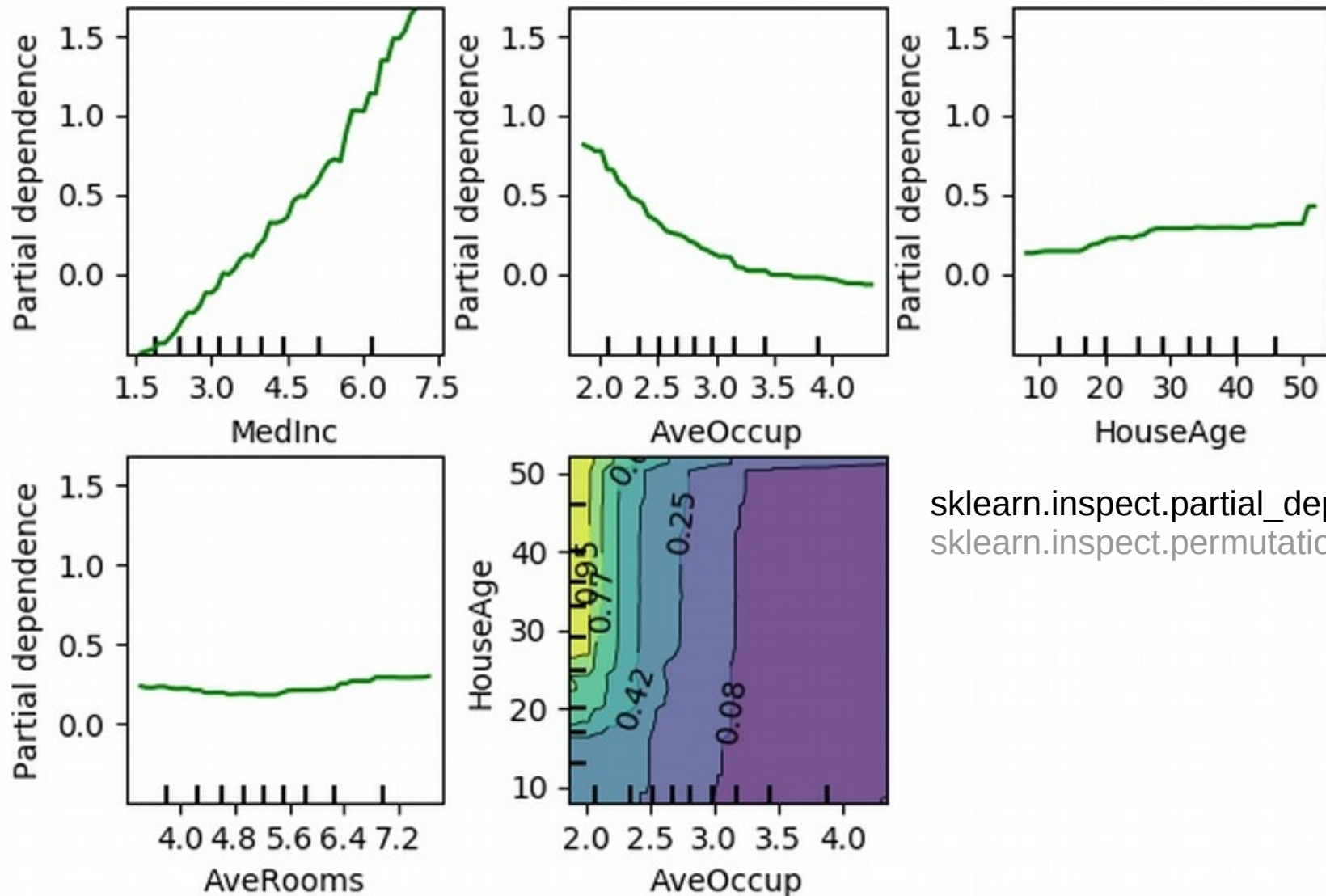
Extreme Gradient Boosting

HistGradientBoostingClassifier



Model understanding and inspection

Partial dependence of house value on non-location features for the California housing dataset, with Gradient Boosting



`sklearn.inspect.partial_dependence`
`sklearn.inspect.permutation_importance`

A Roadmap

<https://scikit-learn.org/dev/roadmap.html>

Governance

<https://scikit-learn.org/dev/governance.html>

Pandas & Feature Names

```
preprocessor = make_column_transformer(  
    (numeric_transformer, numeric_features),  
    (categorical_transformer, categorical_features),  
    remainder='drop')  
  
preprocessor.fit(X_train)  
preprocessor.get_feature_names()
```